REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-03-

0161

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	1/15/03	Final Report covering 5/1/00-10/31/02

4. TITLE AND SUBTITLE
A High Efficiency Algorithm for Combinatorial Synthesis: Maximum Utilization of Minimal Libraries

5. FUNDING NUMBERS
F49620-00-1-0297

6. AUTHOR(S)
Herschel Rabitz

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Princeton University
Department of Chemistry
Princeton, NJ 08544

8. PERFORMING ORGANIZATION REPORT NUMBER
AF 6773-3

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research
AFOSR/NM
801 North Randolph Street, Room 732
Arlington, VA 22203-1977

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release,
distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

This research addressed the common mathematical characteristics of a broad variety of science and engineering applications having large numbers of input variables for exploration. A family of High Dimensional Model Representation (HDMR) techniques was developed to provide efficient and accurate input-output maps for such problems. HDMR operates by breaking down the original high dimensional map into a hierarchy of interrelated lower dimensional components. The mathematical and numerical foundations of HDMR were developed, along with demonstrations of its capabilities in selected chemistry, physics, and engineering applications.

20030513 069

14. SUBJECT TERMS
High dimensional model representation, Functional mapping, Interpolation, Molecular properties

15. NUMBER OF PAGES
6

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

A HIGH EFFICIENCY ALGORITHM FOR COMBINATORIAL SYNTHESIS: MAXIMUM UTILIZATION OF MINIMAL LIBRARIES

F49620-00-1-0297

Herschel Rabitz
Department of Chemistry
Princeton University

Objectives

Many problems in the science and engineering disciplines have very large numbers of input variables. A common necessity is an exploration of these high dimensional spaces, with the effort conventionally thought to grow exponentially with the number of variables in a prohibitive manner. This research developed a special set of high dimensional model representation (HDMR) tools specifically honed for treating such problems. The HDMR technique is capable of reducing the sampling effort from being exponential down to only polynomial scaling in the number of variables. Applications of the concept abound both in mathematical modeling, as well as in the direct treatment of high dimensional field-generated and industrial data sets. The research developed the mathematical and numerical concepts of HDMR, as well as demonstrated its capabilities in selected applications. These applications originally were focused on enhancing the capability of discovering new molecular materials, but they have also been extended to other engineering and scientific domains.

Summary of the Research Effort During the Period of the Grant

During the period of this grant, many of the basic conceptual and mathematical foundations of HDMR were set forth. The underlying concept was to break down a high dimensional input-output problem into a set of interconnected low dimensional subspaces. By using suitable metrics and definitions of orthogonality, representations for the individual low dimensional components may be determined from suitable modeling or laboratory/field data, ultimately resulting in a full HDMR which captures the input-output behavior of the entire high dimensional system. Techniques were developed for sampling the input variable space, ranging from utilizing ordered to randomly located data points. The scaling of these various procedures in terms of effort was explored, with mathematical arguments and evidence indicating that dramatic savings (even up to orders of magnitude in effort) can be achieved with these techniques.

Accomplishments

A. Theoretical and Algorithmic Foundations of High Dimensional Model Representations (HDMR).

(1) High Dimensional Model Representations[1]. An important concern when utilizing HDMR is the number of experiments or modeling excursions necessary to effectively learn the system input $\rightarrow$ output behavior, which is typically a non-linear relationship. Although simple logic suggests that the number of runs could grow exponentially with the number of input variables, broad evidence indicates

that the required effort often scales far more comfortably. This work reviewed the emerging family of HDMR concepts and techniques capable of dealing with such input → output problems in a practical fashion. A summary of the state of the subject was presented.

(2) *Efficient Implementation of High Dimensional Model Representations[2].* The main assumption of HDMR is that for most well defined physical systems the output may be approximated by a sum of hierarchical functions whose dimensionality is much smaller than the dimensionality of the full system. The formulation of HDMR in this work assumed that the data was randomly scattered throughout the domain of the input. It was shown that the number of samples needed for output representation to a given tolerance was invariant to the dimensionality of the input variable space, thereby providing for a very efficient means to perform high dimensional interpolation. Selected applications of HDMR' s were presented from sensitivity analysis and time-series analysis.

(3) *High Dimensional model Representations Generated from Low Dimensional Data Samples I: MP-Cut HDMR[3].* For a high dimensional system, an output $f(\mathbf{x})$ is commonly a function of many input variables $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ with $n \sim 10^2$ or larger. When the approximations given by the first and the second order HDMR correlated functions are not adequate, this work introduced a monomial-based preconditioned HDMR method to represent the higher order terms of a HDMR expansion. The accuracy of the HDMR expansion was significantly improved using pre-conditioning with a minimal number of additional input-output samples.

B. Molecular and Solid State Materials Applications.

(1) *Substituent Ordering and Interpolation in Molecular Library Optimization[4].* The substituent ordering problem in molecular libraries refers to identifying a rational ordering for molecular moieties, such that coarse sampling and interpolation over the full space of possible library molecules may be efficiently performed. This work proposed a practical solution to the ordering problem, based on (a) coarse sampling of the molecular substituents, (b) radial basis function interpolation over the full space, and (c) the use of genetic algorithms to find rational moiety orderings. The algorithm was also used to reorder and predict the glass transition temperature T_g from laboratory data in a combinatorial polymer synthesis library.

(2) *Material Properties Obtained by Using the Correlated Function Expansion for the Quaternary Alloy $Ga_xIn_{1-x}P_yAs_{1-y}$[5].* The band gap and the bond length of the quaternary alloy $Ga_xIn_{1-x}P_yAs_{1-y}$ were estimated over the entire composition space using a HDMR interpolation technique. The HDMR band gap results reproduced well the experimental data for 35 samples of the alloy with an average error of 1.5%. The HDMR band gap variation of $Ga_xIn_{1-x}P_yAs_{1-y}$ lattice matched to InP was in good agreement with existing experimental data. This demonstrated that HDMR can aid in the discovery of new materials with desired properties.

C. Applications to Mathematical Modeling.

(1) *Radiation Transport Simulation by Means of a Fully Equivalent Operational Model[6].* The radiation transport component of atmospheric modeling codes is typically a major contributor to the overall execution time. This research introduced a HDMR as a high-speed replacement for traditional transport modules. The input information to the HDMR was the atmospheric trace gases and temperature profiles as a function of altitude, as well as the surface temperature and albedo. The HDMR was shown to be better than 97% accurate over a broad input variable range, while simultaneously being approximately 10^3 times faster than the traditional radiation transport module it replaced.

(2) *Optimal Control of Catalytic Methanol Conversion to Formaldehyde[7].* An optimal control methodology was applied to find the heat and oxygen flux profiles, distributed along the length of a plug flow reactor, for the conversion of methanol to formaldehyde. A fully non-linear robustness analysis of the formaldehyde production with respect to the catalyst model variables was performed by the use of a HDMR. This representation is similar to the ANOVA decomposition used in statistics but did not require an increase in the number of data points as the dimensionality of the variable space increased.

(3) *Global, Nonlinear Algorithm for Inverting Quantum-Mechanical Observations[8].* This work presented a global, nonlinear HDMR algorithm for extracting molecular potentials from measurements of quantum-mechanical observables. The method utilized a mapping technique to learn the relationship between a broad domain of potentials and their resulting observables to facilitate the inversion. The inversion of laboratory differential cross-section data for He + Ne scattering was performed. The example provided a benchmark for the capabilities of the proposed algorithm to stably reveal the full distribution of potentials consistent with the data.

(4) *Constructing Global Functional Maps Between Molecular Potentials and Quantum Observables[9].* The relationships that connect potential energy surfaces to quantum observables can be complex and nonlinear. In this work, a HDMR approach toward globally representing and exploring potential-observable relationships using a functional mapping procedure was developed. It was demonstrated that an observable's behavior could be learned reliably as a function of the potential and any other variables needed to specify the quantum system.

(5) *Achieving the Laboratory Control of Quantum Dynamics Phenomena Using Nonlinear Functional Maps[10].* This research introduced a HDMR algorithm for achieving closed-loop laboratory control of quantum dynamics phenomena. The procedure made use of nonlinear functional maps to exploit laboratory control data for revealing the relationship between control fields and their effect on the observables of interest. To demonstrate the operation of the proposed map-based

control algorithm, two illustrations involving simulated population transfer experiments were performed.

(6) *Computationally Efficient Atmospheric Chemical Kinetic Modeling by Means of High Dimensional Model Representation (HDMR)[11].* For relieving the computational burden of chemical kinetic calculations in air quality models, this work presented an application of the HDMR method to greatly accelerate these calculations. A HDMR expansion is used to calculate output species concentration at a given reaction time based on the initial input species concentrations. The application of the HDMR method focused on a photochemical box model study of complex alkane/NO_x/O_3 photochemistry. The HDMR expansion was about 400 times faster than the original box-model.

Personnel Supported

Herschel Rabitz, Principal Investigator
Genyuan Li, Research Chemist
Omer Alis, Graduate Student (Ph.D. received, January, 2001)
John Geremia, Graduate Student (Ph.D. received, June, 2001)

New Discoveries, Inventions, or Patent Disclosures

"High Efficiency Mapping of Molecular Variations to Functional Properties"; Patent Application filed January, 2001

"A Machine for Optimal Closed Loop Identification of Molecular Characteristics"; Patent Application, Filed February, 2002.

Publications/References

[1] High Dimensional Model Representations, G. Li, C. Rosenthal, and H. Rabitz, *J. Phys. Chem. A*, **105**, 7765-7777 (2001).

[2] Efficient Implementation of High Dimensional Model Representations, Ö. Alis and H. Rabitz, *J. Math. Chem.*, **29**, 127-142 (2001).

[3] High dimensional model representations generated from low dimensional data samples. I. mp-Cut-HDMR, G. Li, S.-W. Wang, C. Rosenthal, and H. Rabitz, *J. Math. Chem.*, **30**, 1-30 (2001).

[4] Substituent Ordering and Interpolation in Molecular Library Optimization, N. Shenvi, J.M. Geremia, and H. Rabitz, *J. Phys. Chem.*, 2003, in press.

[5] Material Properties Obtained by Using the Correlated Function Expansion for the Quaternary Alloy $Ga_xIn_{1-x}P_yAs_{1-y}$, K. Shim and H. Rabitz, *J. Korean Physical Society*, 2000, **37**, 124-128 (2000).

[6] Radiation transport simulation by means of a fully equivalent operational model, J. Shorter, P. Ip, and H. Rabitz, *Geophys. Res. Lett.*, **27**, 3485-3488 (2000).

[7] Optimal Control of Catalytic Methanol Conversion to Formaldehyde, A. Faliks, R.A. Yetter, C.A. Floudas, S. Bernasek,, M. Fransson, and H. Rabitz, *J. Phys. Chem. A*, **105**, 2099-2105 (2001).

[8] Global, nonlinear algorithm for inverting quantum-mechanical observations, J.M. Geremia and H. Rabitz, *Phys. Rev. A*, **64**, 022710-1-13 (2001).

[9] Constructing global functional maps between molecular potentials and quantum observables, J.M. Geremia, H. Rabitz, and C. Rosenthal, *J. Chem. Phys.*, **114**, 9325-9336 (2001).

[10] Achieving the Laboratory Control of Quantum Dynamics Phenomena Using Nonlinear Functional Maps, J.M. Geremia, E. Weiss, and H. Rabitz, *Chem. Phys.*, **267**, 209-222 (2001).

[11] Computationally Efficient Atmospheric Chemical Kinetic Modeling by Means of High Dimensional Model Representation (HDMR), S.W. Wang, P.G. Georgopoulos, G. Li, and H. Rabitz, *J. Geophys. Res.*, 2003, in press.